Versatile Diffusion: Text, Images and Variations All in One Diffusion Model

Year: 2022 | Citations: 238 | Authors: Xingqian Xu, Zhangyang Wang, Eric Zhang, Kai Wang, Humphrey Shi

---

Abstract

Recent advances in diffusion models have set an impressive milestone in many generation tasks, and trending works such as DALL-E2, Imagen, and Stable Diffusion have attracted great interest. Despite the rapid landscape changes, recent new approaches focus on extensions and performance rather than capacity, thus requiring separate models for separate tasks. In this work, we expand the existing single-flow diffusion pipeline into a multi-task multimodal network, dubbed Versatile Diffusion (VD), that handles multiple flows of text-to-image, image-to-text, and variations in one unified model. The pipeline design of VD instantiates a unified multi-flow diffusion framework, consisting of sharable and swappable layer modules that enable the crossmodal generality beyond images and text. Through extensive experiments, we demonstrate that VD successfully achieves the following: a) VD outperforms the baseline approaches and handles all its base tasks with competitive quality; b) VD enables novel extensions such as disentanglement of style and semantics, dual- and multi-context blending, etc.; c) The success of our multi-flow multimodal framework over images and text may inspire further diffusion-based universal AI research. Our code and models are open-sourced at https://github.com/SHI-Labs/Versatile-Diffusion.